DOCUMENT RESUME

ED 278 265 FL 016 404

AUTHOR	Clark, John L. D.
TITLE	Handbook for the Development of Tape-Mediated, ACTFL/ILR Scale-Based Tests of Speaking Proficiency in the Less Commonly Taught Languages.
INSTITUTION	Center for Applied Linguistics, Washington, D.C.
SPONS AGENCY	Department of Education, Washington, DC.
PUB DATE	Dec 86
GRANT	G008402258
NOTE	27p.; For related document, see FL 016 403.
PUB TYPE	Guides - Classroom Use - Guides (For Teachers) (052)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Audiotape Recordings; Chinese; Communicative Competence (Languages); Correlation; Evaluation Methods; *Language Proficiency; *Language Tests; Models; Speech Skills; *Test Construction; Test Format; Test Use; Test Validity; *Uncommonly Taught Languages; *Verbal Tests
IDENTIFIERS	ACTFL ILR Oral Proficiency Guidelines

ABSTRACT

This handbook resulted from a project to develop a tape-based alternative test of language proficiency in less commonly taught languages. The objective was to produce a test that is modeled on and readily interpretable by the American Council on the Teaching of Foreign Languages/Interagency Language Roundtable (ACTFL/ILR) proficiency guidelines and that can serve as an alternative to the live interview in situations where the interview is not financially or administratively feasible. The handbook documents and provides commentary on the development and validation activities carried out in producing a tape-based test for Chinese. Sufficient depth and detail are provided to serve as a general guide for the preparation of similar tests in other less commonly taught languages. The conceptual and practical questions addressed in the handbook may be applicable to other languages even though different languages may present somewhat different measurement needs. A questionnaire for participants in the Chinese Speaking Test Study is appended. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED278265

Handbook for the Development of Tape-Mediated, ACTFL/ILR Scale-Based Tests of Speaking Proficiency in the Less Commonly Taught Languages

John L. D. Clark

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C.R. Tucker

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Center for Applied Linguistics
1118 22nd Street, NW
Washington, DC 20037

December 1986

FL016404

BEST COPY AVAILABLE

OVERVIEW

During 1984-1986, with funding support from the U.S. Department of Education, the Center for Applied Linguistics (CAL) developed and validated a series of tests of speaking proficiency in Chinese that make use of a "semi-direct" procedure. This is a procedure in which the test stimuli are presented to the examinee through a tape recording and associated test booklet, rather than through direct face-to-face conversation as is the case with the ACTFL/ILR oral proficiency interview (OPI). Although the OPI procedure has been used extensively and with good success for several of the languages most widely taught in the United States (e.g., French, Spanish, German), the relatively small student enrollments in the so-called "less-commonly-taught" (LCT) languages--together with the associated limitations on material and personnel resources within these language areas--may often make it difficult or impossible to establish the necessary infrastructure of highly trained interviewer/raters to conduct interview-based testing or, by the same token, to establish and maintain a suitable mechanism for the monitoring and quality-control of the testing process.

It was the intent of the CAL project to develop an alternative approach to speaking proficiency testing that would be operationally practical for use in an LCT language setting, but at the same time highly congruent with the regular ACTFL/ILR interviewing and rating procedure in terms of both the types and extent of language use elicited in the course of the test and the nature and interpretation of the testing results. In keeping with these goals, the project developed four alternate forms of a tape- and booklet- mediated Chinese speaking test, designed to be administered either in a student recording-capable language laboratory or in a regular classroom using two portable tape recorders, one to play the master test tape and the other to record the student's responses. The four test forms were subsequently validated by administering both the project tests and a live face-to-face interview to the same groups of students and determining the degree of correspondence of the scores obtained on both types of instruments. The statistical indications from this study are that, with properly trained raters, the semi-direct tests can be administered and scored as reliably as the face-to-face interview, and on this basis can be used as effective surrogates for the live interview in situations where face-to-face testing is not financially or administratively feasible. (A detailed description of the validation study and its results are available in the final project report, "Development, Validation, and Dissemination of a Proficiency-Based Test of Speaking Ability in Chinese and an Associated Assessment Model for Other Less Commonly Taught Languages" [U.S. Department of Education Grant No. G008402258], 1986, available from the Center for Applied Linguistics.)

The purpose of the present handbook is to document and provide commentary on the development/validation activities carried out for the Chinese speaking tests in sufficient depth and detail to serve as a useful general guide for the preparation of similar tests in other less-commonly-taught languages. In this regard, it should immediately be acknowledged that the implementation of a similar test development project within a different language or language family

may quite properly give rise to somewhat different measurement needs and solutions from those represented by the Chinese project. However, it is felt that documentation of the conceptual questions wrestled with and practical insights gained in the course of the Chinese testing work may be of some assistance to persons considering the development of this type of test in other languages.

DIRECT VS. SEMI-DIRECT TESTING

As defined and discussed in detail by Clark (1979), a "direct" testing procedure is one in which a human tester interacts on a real-time basis with a given examinee, both to present the test stimuli and (in most instances) simultaneously evaluate the examinee's responses. "Semi-direct" tests, by contrast, are defined as any procedures which elicit active speech on the part of the examinee by means of printed booklets, tape recorded stimuli, or other "non-human" means, rather than through direct interaction with a live interlocutor.

A major advantage of the direct test--for which the oral proficiency interview (OPI) is the primary example--is the high degree of face validity and procedural validity which it embodies. As a means of determining speaking competence in a conversational setting, it is difficult to conceive of a more realistic testing approach than actually engaging the examinee in conversation. The interactive nature of live conversation permits the demonstration, on the examine's part, of the ability or lack of ability to use appropriately a variety of discourse management strategies of major importance to real-life conversation, including following appropriate turn-taking conventions, requesting clarification as necessary, repairing miscommunications, etc.

In considering the use of a semi-direct test as a measure of speaking proficiency, it must be acknowledged from the outset that interactive, discourse-management aspects of the student's overall speaking proficiency cannot readily be elicited (or by the same token, effectively measured) through semi-direct techniques. However, this limitation notwithstanding, it is felt that certain types of semi-direct techniques, which for convenience can be referred to as proficiency-oriented semi-direct procedures, can serve to validly and efficiently measure many of the other performance aspects that constitute overall speaking proficiency as this concept is understood and applied in the ACTFL/ILR interviewing/rating context. Here and throughout the workbook, the term "proficiency-oriented semi-direct test" will be used to refer to any measurement procedure that (1) approximates as closely as possible the linguistic content and manner of operation of a live interview, and (2) is scored and interpreted on the basis of the ACTFL/ILR proficiency guidelines or other functionally similar descriptive scale. The Chinese Speaking Test (CST) was planned and developed with these criteria constantly in mind, and the numerous discussions that were held concerning the overall format for the test as well as the individual question types to be employed all had as crucial touchstones the two interrelated questions: "How can we best approximate the elicitation procedures and contextual conditions of a well-conducted live interview through the vehicle of a a stimulus tape and booklet?" and "To what extent will a proposed format or question type tend to elicit a student response that can be efficiently and reliably evaluated on the basis of the ACTFL/ILR scale?"

A MAJOR CAUTION

The proper interpretation and effective use of the material contained in this handbook presupposes an extensive theoretical and practical background--as well as considerable "hands-on" experience--in oral proficiency interviewing and rating as both of these areas have been elaborated by the American Council on the Teaching of Foreign Languages (ACTFL) in conjunction with Educational Testing Service (ETS), the Interagency Language Roundtable (ILR), the Center for Applied Linguistics (CAL) and other organizations that have played a major role in the development and dissemination of this measurement technique. Although the information provided in this workbook is intended to be as detailed as possible within necessary scope and length constraints, study of this material CANNOT in and of itself provide the necessary background and familiarity with oral proficiency testing concepts and procedures in general that would be needed to properly interpret and make effective use of the information presented. Persons intending to make use of this workbook as an aid to test development should, at a minimum, have successfully completed a full-scale ACTFL-sponsored interviewer/rater training workshop either in the language of the test or, where necessary, a similar ACTFL-led workshop using English-medium presentations and example interviews. All of the following sections of the handbook will presume at least this degree of familiarity and experience with the oral proficiency interviewing/rating process and its conceptual foundations.

OVERALL TEST FORMAT: SOME BASIC CONSIDERATIONS

Seven basic considerations governed the specification of the overall testing format for the CST, and these same issues are considered relevant to the design of similar tests for other languages.

1. To provide maximum flexibility in the types of stimuli that can be presented to the student, the test should make use of a printed booklet as well as a master audiotape.

Prior to the CST, there was only one known test development effort aimed at producing a "proficiency-oriented semi-direct test" as the term is used here. This was the Recorded Oral Proficiency Examination (ROPE), designed and reported by Lowe and Clifford (1980). The ROPE consists of a series of tape-recorded questions in the target language, which the student is to answer orally one-by-one during appropriately timed pauses on the tape. Although this question-answer format does provide a reasonable amount of flexibility, in that questions across a variety of topical areas and at varying levels of difficulty can be posed, it has several drawbacks. First, especially for the more sophisticated questions, it is possible that examinees might not fully understand the nature or intent of the orally-presented target language question, even though they might have sufficient speaking proficiency to respond appropriately to the identical question presented in some other manner (e.g., printed in a test booklet). In the live oral interview, the mis-hearing or misunderstanding of a spoken question is not as great a problem, since the examinee can ask for repetition, the question can be rephrased in a simpler form, etc. However, in the tape-recorded mode, there is no opportunity for

this type of clarification, and miscomprehension of a spoken question can, in the absence of visual reinforcement in a test booklet, inappropriately negatively affect the student's active response to the tape-recorded question.

A second and even more compelling reason to combine a printed booklet with a master audiotape was to provide for greater flexibility in eliciting extended descriptions, narrative discourse, and other types of examinee speech in addition to conversational question-answer per se than would be possible in a tape-only format. In the live interview, face-to-face conversation is often supplemented by the use of a variety of "visuals," including, for example, index cards on which are printed short descriptions of a particular communicative task (e.g., renting a hotel room, having an article of clothing repaired, etc.) which the student is to carry out based on the description provided. Use of a test booklet along with the tape makes it possible not only to provide these "situation" or "role-play" stimuli but also present a variety of other printed or pictorial material intended to elicit other types of responses (such as extended narrative description) that are at issue in the level descriptions.

2. The test should cover the level 1 to 3+ range (Intermediate Low to Superior) on the ACTFL/ILR scale.

With respect to the "higher bound" of the intended test range, the selection of 3+ was, as a practical matter, mandated by the semi-direct nature of the test. At level 4 and above, the examinee's proficiency is such that detailed and explicit probing on an interactive (i.e., live conversation) basis is necessary to adequately determine the extent and depth of lexical knowledge, ability to respond appropriately to abrupt conversational turns, and a variety of other higher-level performances at issue in the "4-and-above" range. At the other end of the scale, a "lower bound" of level 1 was selected for two reasons. First, in order to present test stimuli appropriate for the 0+ student, it would be necessary to make use of a variety of rather artifical procedures (e.g., naming colors, identifying objects in a room, etc.) that would not ordinarily be encountered in a real-life communication setting and, as such, would be very much out of keeping with the more realistic and more face- and content-valid communicative tasks at issue elsewhere in the test. Second, it was felt that to include provisions for eliciting "below level 1" types of memorized or rote responses within the body of the test might serve to inappropriately legitimize these types of responses as representative of some acceptable level of real-life functional capacity in the language.

3. Each of the questions in the test should be answerable, in at least a minimal fashion, by examinees at the lowest end (level 1) of the score range.

Because the taped test would necessarily be administered in a linear fashion, with all examinees, regardless of proficiency level, attempting each of the questions, it was considered important to insure that even the least competent students would be able to say at least something in response to each question. Both in the question planning and writing/reviewing process, an important touchstone was to make sure that even a "survival level" examinee would not be completely blocked by any of the test questions. For example, even in the most sophisticated "situation" tasks, e.g., making appropriate

after-dinner remarks in a highly formal setting, even level 1 examinees would be able to say the target language equivalent of "Thank you, like very much, hope come again," etc.

4. Wherever possible, each question should provide the opportunity for more sophisticated and elaborated responses on the part of higher-level examinees.

It was acknowledged that certain questions on the test would necessarily fail to elicit a full level 1 - 3+ range of responses because the topical area and/or the communicative requirement was intrinsically at a lower level. For example, in the conversational portion of the test, questions such as "Where do you live?" would not be expected to probe beyond level 2, since this is not a situation in which even educated native speakers would normally produce a "beyond level 2" type of response. However, to the extent possible, each question in the test was phrased so as to include some type of cue that would alert the more proficient examinee to the desirability of making a more elaborated response. For example, one question asked the examinee to describe the typical day of an American office worker, based on a sequence of line drawings. The additional instruction, "You may wish to add some appropriate social commentary," was intended to elicit some discussion of the overall American work ethic, the fast-paced "lunch at your desk" syndrome, etc. which was clearly implied in the drawings and would be a natural topic for comment by examinees having the requisite level of proficiency.

5. The test should follow the normal progression of the live interview with respect to both the kinds of questions asked and the level of sophistication targeted by these questions.

Just as the live interview is intended to progress from eliciting relatively simple factual information in an autobiographical mode up through progressively more sophisticated types of language performance, the CST was designed to follow a similar general sequence of autobiographical information, narrative descriptions, past-time and future-time descriptions, responses requiring comparisons, supported opinion, discussion of hypothetical situations, etc., up through a final series of role-playing situations requiring tailoring of content and style of speech in keeping with the sociolinguistic characteristics of each situation.

6. With the single exception of the target language questions in the "live conversation" section of the test, all test directions and test stimuli should be in English.

This consideration was definitely not in keeping with conventional or traditional approaches to testing within the Chinese field, in which the use of English is characteristically avoided except possibly for the initial general (whole test) directions. However, it was felt that presenting the test directions or the individual question stimuli in Chinese would run the substantial risk of having the examinee not understand (or only partially or erroneously understand) the specific speaking task at issue and thus fare poorly on the test--the actual level of speaking proficiency notwithstanding.

For the higher-level questions especially, where it would be necessary to convey quite complex and/or subtle information to the examinee in order to properly communicate the intended nature of a response, limitations on the examinee's listening comprehension or reading proficiency in Chinese might mean that these complexities and distinctions would be lost to the examinee, and thus negatively affect or completely block an appropriate response, regardless of his or her level of speaking proficiency per se. By providing the test instructions and other stimuli in English (with the single exception of the conversational section, in which this approach would be quite artificial and inappropriate on face validity grounds)--and also by arranging to provide most of these instructions in written as well as spoken form--the student's chances of not understanding or misunderstanding the required response tasks as such were appreciably reduced.

7. Each of the test questions should be framed so as to reflect a realistic and meaningful communicative task within a real-life context.

Throughout the test, each question or question section would be set within a "scenario" that would involve the examinee as a transmitter of novel information to a specified interlocutor for a pragmatic communicative purpose. For example, in the first (conversational) part of the test, instead of simply asking the examinee to "answer the following questions aloud," the instructions were: "...you will have a conversation with a native speaker of Chinese. This person would like to find out some things about your background, activities, and interests. For each question asked by the Chinese speaker, listen carefully and answer during the pause immediately following the question." In a later section eliciting fairly detailed descriptions based on complex line drawings, the instructions were not a linguistically artificial "name everything in the picture" or "say everything you can about the picture," but "A Chinese-speaking friend has expressed an interest in taking her children to a recreation park in your neighborhood this coming weekend. She is very interested in knowing what facilities are available in the park and, in general, what kinds of things there are to do there. You will have 15 seconds to study the picture. Then, when the signal sounds, tell your friend as much as possible about the park and what one can do there." This "scenario" approach to the framing of the test questions was considered to provide a more communicatively natural framework for the examinee's response by indicating clearly the pragmatic purpose for the response, the nature of the "target audience" for the response, and other sociolinguistically relevant information. For higher-level examinees especially, the "scenario" approach also served to alert them to the speech style and discourse level appropriate to a particular question or test section.

OVERALL TEST INSTRUCTIONS

The general instructions for the test, which are delivered by the master voice on the tape as well as printed in the test booklet, are shown in Appendix A. Important points stressed in these instructions include:

1. The range of the test is intentionally broad, and no one is expected to answer all questions completely.

2. The examinee should be sure to speak as much and as well as possible in response to each question.

Also included are reminders to speak loudly enough and clearly enough to produce a good quality recording, and to make sure that the examinee's tape recorder is in the "record" mode, with the tape moving through the machine, before the test begins.

Each of the item-type sections included in the final version of the Chinese test is described in detail below, followed by discussion of rationale, cautions/sugggestions in question preparation, important information to be conveyed in the directions for that section, and so forth. Examples are drawn as required from any of the four versions of the test, each of which is reproduced in full, together with master tape scripts, in Appendix A.

PERSONAL CONVERSATION

In this section, the examinee listens, with test booklet closed, to spoken questions in Chinese delivered by the master voice on the test tape, and responds to each question as it is asked. The scenario established for this section is that "you will have a conversation with a native speaker of Chinese. This person would like to find out some things about your background, activities, and interests." The examinee is advised "For each question asked by the Chinese speaker, listen carefully and answer during the pause immediately following the question. Say as much as you can in response to each question." [Note: The caution to say as much as possible was also given in the general instructions, but is repeated here as an important reminder.] A short "beep signal" (500 Hz, approximately 1/2 second) is given at the end of the allotted response pause as an indication that the next Chinese question is about to be spoken.

- Approximately 12 questions are asked in this section, beginning with very simple questions (target language equivalent of "Hello, how are you?") and progressing to other conversational questions that reach to approximately level 2/2+. The questions are typical of polite conversation among relative strangers (as is implicitly the case in the early portions of the face-to-face interview). Questions broached can include simple biographical information, descriptions of routine activities, narrations in past or future time, comparisons, supported opinion, and other types of language use within roughly the level 1-2+ range.

- The spoken questions should be phrased in a natural and idiomatic manner, but should use very common lexicon and structure, so that the examinee will have little or no difficulty understanding the meaning/intent of each question.

- Each question should be applicable to the actual biographical experience of any examinee who would reasonably be expected to take the test at any time. For example, a question such as "Are you enjoying your studies?" would not be appropriate unless both immediate and possible future administrations were to be restricted to then-current students. Alternatively stated, each question in this section should have "present truth-value" for the examinee. Although this requirement appreciably narrows the scope of possible questions (by comparison

to the live interview, which can be adapted on a real-time basis to the particular background of the examinee), some care in phrasing can help to insure adequate generalizability. For example, instead of asking "Where is your house located?" (presupposing a type of lodging that many student examinees might not possess), the question could be phrased "Where are you living at present?," which would embrace any and all types of housing.

• Conversational rejoinders, if any, made by the tape in "response" to the examinee's preceding answer must be a reasonable, or at least not illogical, reply to whatever the student may in fact have said. For example, to a "Hello, how are you?" question, a taped response of "That's good" would not be an appropriate conversational follow-up for an examinee who had just complained about a splitting headache or who had expressed great anxiety about taking the test. In general, because of the difficulty in judging ahead of time the variety of examinee responses that might be made to a particular taped question, conversational "rejoinders," although frequent in live conversation, are probably best avoided in a tape-recorded context or employed with extreme care. Although an occasional intentionally vague and "all-purpose" reply such as the target language equivalent of "I see" (followed by the next question) could provide for some conversational naturalness, this is probably about the limit of the feasible in a semi-direct mode.

• Careful response pause timing is critical if the "conversation" is to be reasonably natural. In the Chinese versions, typical pauses ranged from 6-7 seconds for the initial short questions ("Hello, how are you?") to approximately 20 seconds for the more discursive questions involving narrative descriptions or comparisons (e.g., "Do you like to eat restaurant food or home cooked meals? Why?"). If possible, a clinical tryout aimed at, among other things, determining the appropriate pause lengths, should be undertaken. This can be done by recording each of the conversational questions end-to-end (without pauses) on a trial tape and--on a one-to-one basis with 8-10 examinees typical of the full proficiency range envisioned for the test population--starting and stopping the tape as necessary for each examinee, and recording the response times actually used. Experience with the Chinese test suggests that, for any given question, response times near the high end of the observed range are preferable to shorter pauses since examinees, for the most part, are more frustrated by being "cut off" during a reply than by having some seconds of waiting time prior to the next question.

PICTURE-BASED DISCOURSE

This section is characterized by the presentation, in the test booklet, of both single line drawings and sets of 4-5 topically-related "picture sequences." Both the individual pictures and the picture sequences are accompanied by spoken English questions on the master tape. In the Chinese tests, two single-picture items were included in each test version, with the first of these intended to check the examinee's ability to give simple directions. As in all parts of the test, the picture-based tasks are set within genuine communicative "scenarios." For example, the "giving directions" question for one version of the test was phrased as follows: "Imagine that you have invited a Chinese friend to come to dinner at your house, which is a few miles out of town. You need to give your friend directions over the telephone. These directions are shown by a dotted line on the map. You will have 15

seconds to study the map. Then, when you hear a tone signal, tell your friend how to drive from his house (which is marked by an X at the bottom right) to your house."

The second single-picture question for each test version was aimed at checking the examinee's ability to provide detailed descriptions, and included a rather complex scene involving both common and higher-level lexicon. A typical scenario: "A Chinese-speaking friend has expressed an interest in taking her children to a recreation park in your neighborhood this coming weekend. She is very interested in knowing what facilities are available in the park and, in general, what kinds of things there are to do there. You will have 15 seconds to study the picture. Then, when the signal sounds, tell your friend as much as possible about the park and what one can do there."

With regard to the picture-sequence items, the final version of each of the Chinese tests included three sequences--the first intended to elicit descriptions of habitual activities (in present time) the second, a past time narration, and the third, a future time narration. Typical scenarios included:

(Habitual action:)

A typical day in the life of an average American office worker.

A typical school day for an American elementary school teacher.

A typical Saturday for an American family.

(Past time narration:)

A traffic accident.

A series of unfortunate events during a hotel stay.

Transportation problems in trying to get to a meeting.

(Future time narration:)

Plans for tourism and shopping during the coming weekend.

Plans for things to do with a Chinese friend visiting the United States.

Plans for a backpacking trip in the countryside.

For both the single-picture and picture-sequence drawings, the style, degree of detail, and other visual aspects of the drawings are crucial to the measurement validity of the test questions. The following points should be noted:

• Drawings are preferable to photographs in that (1) photographs usually contain extraneous and confusing detail that can often "hide" the particular

aspects of the scene that are most at issue in the test; and (2) if photographs are used, the test developer may be tempted to test whatever kinds of things happen to be exemplified in available photographs, rather than carefully specifying ahead of time the intended linguistic situations and anticipated student responses, and tailoring the test stimuli to meet these requirements.

• Since both men and women will be taking the test, the "protagonist" in a given picture sequence (if intended to represent the examinee) should be drawn as a unisex character to the greatest extent possible. Slacks/trousers (rather than dresses or obviously male suits), hair of a length that could be worn by either sex, etc., will all contribute to this effect. By the same token, the situation itself should be one in which both men and women would be equally likely to find themselves. [Note: These considerations apply only to those situations in which the examinee him/herself is represented in the drawings and a first-person narration is at issue. When third-person narration is intended, the "main character" in the drawings may be obviously of either sex.]

• A modified cartoon style, in which the people are shown in a simplified and stylized (but not "funny") manner, is preferable to more realistic and more complex depictions. The scene itself should also be highly stylized and simplified to the bare essentials necessary to convey the intended information.

• For drawings intended to elicit "giving directions," experience with the Chinese tests suggests that an oblique view of an actual scene (cf. p. 1, Form D) is preferable to either a genuine map or a stylized bird's-eye map). The reason for this is that both the genuine map and, to a lesser degree, the bird's-eye drawing put a premium on the examinee's ability to "read maps"--an extraneous factor insofar as target language proficiency per se is concerned. The oblique perspective provides a "mind's eye" view of the intended route and frees the examinee from the potentially confounding chore of map reading as such.

• For drawings intended as stimului for detailed descriptions, enough objects or activities should be depicted to allow for an elaborated response on the part of more proficient examinees, without completely frustrating the less competent examinees. For example, in Form A, p. 2, the neighborhood park scene includes a variety of linguistically sophisticated items from a learner's perspective (snack cart, children going down a slide, sand box, etc.) that could be precisely related or at least readily paraphrased by a higher-level student, but at the same time also includes other easier items ("swimming," "playing tennis") that would be within the ability of a less proficient examinee.

Close and effective communication between the test developer and artist is needed to insure that the drawings will be of the highest measurement value. A suggested approach is as follows:

(1) The artist should be informed in detailed but "lay" terms about the nature of the test, the age and background of the anticipated examinees, the role of the line drawings in eliciting certain types of responses on the examinees' part, etc. If possible, the artist should sit in on a session in which several of the picture-based questions are discussed and planned.

(2) For each individual picture or picture sequence, a detailed information sheet should be prepared by the test developer, including:

- A brief description of the linguistic situation or task facing the examinee in that particular instance.

- A verbal description of the desired content of and/or actions to be depicted in each picture, emphasizing the major elements (from a student response perspective) at issue.

- If possible, a very rough free-hand sketch of each picture (especially for the picture sequences) to show the general layout and orientation. This recommendation presupposes a good degree of visualization ability and some artistic competence on the part of the test developer; where this is not the case, test developer-produced sketches may be counterproductive. In all instances, the artist should be explicitly urged to suggest improvements, revisions, or even completely new approaches to any visual outlines provided by the test developer.

(3) Penciled working drafts should be provided by the artist and reviewed carefully by the test developer before the final inked drawings are prepared. Also at this stage, a clinical tryout should be conducted by asking a number of native speakers to answer the intended questions on the basis of the drawings provided. For each picture or picture sequence, native speakers should be readily able, with a quick look at the pictures, to "follow the script" very closely in terms of the items mentioned, the events described, etc. Any hesitation or confusion, misinterpretation of the scene, etc. will indicate the need for revision of the drawings.

In the Chinese tests, for both the single picture and picture-sequence sections, a recurring question was the degree of emphasis to be accorded to U.S.-based settings by comparison to Chinese settings. It was acknowledged that, from a strictly viewed "proficiency" standpoint, examinees at a given score level should be expected to be able to speak equally well about places, activities, and events within both U.S. and Chinese settings. However, depending on the particulr type of teaching program that the examinee has undergone, he or she may not have been exposed to any extent to the lexical items (or other pertinent linguistic aspects) at issue in one or the other of these cultural contexts. For example, students using text materials based on mainland China or Taiwanese settings may not have been exposed to even relatively simple food, clothing, etc. vocabulary as applicable to a Western context. In an attempt to minimize the potential negative effects on examinees whose language learning experiences have been exclusively or predominantly in a single cultural setting--while remaining as faithful as possible to the concept of generalized, non-curriculum specific, proficiency testing--the following guidelines were followed in developing the picture-based sections (as well as the other test sections):

● To the extent possible, a rough balance should be made between scenes situated in the U.S. and in a Chinese country. For example, in Form B, the "giving directions" task is set in the Chinese countryside; the "detailed description," in an American supermarket; and the present, past, and future narrations, respectively, cover a day in the life of an American schoolteacher.

problems encountered in a New York hotel, and plans for shopping and tourism during an upcoming visit to Beijing.

● Notwithstanding the formal location of a given scene, the particular elements to be dealt with by the examinee should be culturally neutral and equally possible in both geographic settings. For example, in the New York hotel setting, most of the problems depicted in the drawings (insufficient water supply, noise from an adjoining room, waiter-spilled food) could certainly be encountered in either U.S. or Chinese contexts.

RESPONSES TO PRINTED QUESTIONS

For each question in this part, the examinee reads a printed question in English (which is also voiced on the tape) and answers it during the pause provided. The primary measurement objective of this part of the test is to provide for an extension of the first ("conversation") part of the test into areas of discourse that go beyond simple question-answer exchanges. These more extended performances, which collectively are aimed at probing aspects of proficiency at issue between roughly levels 2 and 3/3+, include the following types of examinee performances.

(Extended descriptions of personal interests and activities:)

"Please describe the kinds of cultural activities you particularly enjoy."

"Please talk about what you like to do for recreation."

"Please talk about the kinds of things you like to read for pleasure."

(Comparisons:)

"Talk about the advantages and disadvantages of being a student (as opposed to working full time)."

"Which do you prefer, Chinese or American food, and why?"

"Talk about the advantages and disadvantages of having a television set in your home."

(Complex descriptions:)

"Explain to a Chinese friend how credit cards are used in the United States."

"Explain to a Chinese friend how one would go about buying a used car in the United States."

"Explain to a Chinese friend how one would go about renting an apartment in the United States."

(Supported opinion:)

"Some people feel that the primary goal of education should be to prepare students to better compete in the job market. Do you agree or disagree? Explain why."

"Some people feel that the influx of people from foreign countries is making it difficult for American citizens to get work. Others argue that the strength of America lies in its cultural diversity, and that the country benefits from the presence of foreign-born persons. With which of these two positions would you most strongly agree? Give reasons for your opinion."

(Hypothetical situations:)

"Suppose you were asked to defer your current career plans for a period of one or two years and to carry out some other activity of direct and immediate benefit to your own or another society. Describe what you might do and how this might affect society."

"Suppose that an acquaintance is considering studying Chinese and has come to you for advice and guidance about the best approach to follow. What would your recommendations be?"

"If it were financially unnecessary for you to continue working beyond age 45, what do you think you would want to do over the course of the rest of your life? Why would you choose these particular activities?"

Question-preparation suggestions and additional discussion of the "responses to printed questions" item type are given immediately below:

- As in the simulated conversation section, any question posed in this section should be applicable to any potential examinee. For example, the question asking for "advice and guidance" concerning Chinese study presupposes that the examinee has formally studied Chinese in the past and/or has had other kinds of experiences that would provide a reasonable knowledge background in this area.

- The questions should deal with topics that are about equally familiar to male and female examinees. For example, a general question about food preferences ("Which do you prefer, Chinese or American food, and why?") would be appropriate for both male and female examinees, whereas a question dealing with particular recipes, cooking techniques, etc. might not.

- For the "complex description" questions, each topic should be one that a native English speaker might legitimately have occasion to talk about with a Chinese-speaking interlocutor in a real-life conversational situation. In this regard, explaining "how credit cards are used in the United States" is a more appropriate topic of discourse than, for example, describing banking procedures in Taiwan.

- Prospective "supported opinion" questions should undergo each of the following checks:

- The topic on which an opinion is requested should be one that virtually every examinee would have had some contact with through the the media, general reading habits, etc. For example, "education for careers" vs. "education for personal development" is a "person-in-the-street" topic that virtually any examinee would be expected to have encountered and to have some general thoughts about.

- Topics that are highly controversial or that address potentially sensitive religious, moral, or emotional areas should be avoided, including, for example, abortion, use of heroic life-sustaining measures, the "right" to commit suicide under certain circumstances, etc. Although these matters are indeed frequently discussed in real life, the potential "public relations" consequences of including items of this type in a widely distributed test--as well as the very real possibility of offending or taking aback a number of individual examinees--would argue rather strongly against their inclusion.

- The question itself should be carefully framed, so that the specific area of debate/choice at issue is very clear, and at least the broad outlines of potential arguments that might be made (on both sides) are provided. For example, instead of the quite vague and undirected "What is your opinion about foreigners coming to the United States, and why?," the question actually used in this topical area stated two alternative positions --"Some people feel...difficult for American citizens to get work"; counterposed by "others argue...cultural diversity." The second formulation, while providing a good amount of latitude for the more proficient examinee, also gives needed support to the level 1/1+ respondent in the form of broad-outline suggestions about how a response might be framed.

● A major goal of the hypothetical questions is to provide an opportunity for the examinee to demonstrate his or her capabilities in discussing concepts, issues, and ideas, as distinguished from the more factual topics addressed, for example, in the complex description questions. Two considerations in the preparation of the hypothetical questions are as follows:

- To engage the examinee directly in the hypothesized situation, a "first person" response is asked for: "Suppose that you were asked to defer..."; "If it were financially unnecessary for you to continue working...."

- As is the case throughout the test, potentially emotion-laden situations, e.g., "Suppose that you became very ill," "Suppose you suffered a family tragedy," etc., are avoided.

- Notwithstanding the preceding, the hypothetical situation established should require some serious introspection on the examinee's part, at the level of at least temporary changes in life-style or direction.

- Situations requiring the examinee to assume impossible roles or personae ("Suppose you were the President of the United States..."; "Suppose that you had just discovered America...") should be avoided, both on "real-life communication" grounds and on the grounds that examinees would be expected to have widely differing capabilities to "project" and "pretend" in fanciful situations of this type, regardless of their degree of competence in the target language per se.

SITUATIONS

The last section of the test was intended to correspond, to the greatest extent possible, to the "role-play" activities frequently included at or near the end of the live interview. As such, they were considered to offer the greatest possibility for determining the examinee's ability to select and use the specific lexicon, register, and speech style most appropriate to that particular type of situation and, in general, to demonstrate situationally-determined sociolinguistic competencies not readily elicited in other sections of the test.

The general instructions for this section read in part as follows: "For each situation, you are asked to pretend that you are actually in the situation, and to respond in a linguistically appropriate manner....When you respond, be sure to use the type and style of language that a native Chinese speaker would use in this situation." The following is a typical "situation" question:

> You are standing in line at a train station ticket window in Taipei. There are two people ahead of you and your train is scheduled to leave in five minutes. Ask the people in front if they could possibly let you get ahead of them, explaining the reason for your request.

To be noted in the above are specific or clearly implicit characterizations of (a) the physical setting in which the communication is taking place (train station ticket window in Taipei); (b) the interlocutor(s) with whom the examinee will need to communicate (Taiwanese fellow-ticket purchasers); (c) any temporal or other conditions potentially influencing the nature or extent of the communication (five minutes left before train time); and (d) the style or tone of language most appropriate in this context (although not explicitly stated, the "possibly let you get ahead...explaining the reason" phrase strongly implies an apologetic [and rapid] invocation of external extenuating circumstances).

Across the five situations presented in each test version, there is a progressive increase in both the linguistic competence and psycholinguistic/cultural sophistication required to deal appropriately and effectively with the situation involved. The earliest questions in this section deal with fairly "standard" situations in which affectively unmarked and largely conventional communicative tasks are at issue, for example:

> You are at the front desk of a small hotel on the east coast of Taiwan. Tell the clerk you need a single room with air conditioning and private bath. Inquire about the room rate and ask if you can pay by travelers check.

Later situations involve more sophisticated language use in which some degree of persuasive or argumentative discourse and/or culturally appropriate display of anger, dissatisfaction, or other personal emotion is required:

> You have just finished eating lunch in a crowded, busy restaurant in Kaohsiung. In looking over the check, you find that you have

inadvertently been charged for an additional dish you did not order, and also that the price of one of the dishes you did order is higher than was shown on the menu. Call the waiter and describe these discrepancies. Ask for the bill to be corrected.

Several days ago, you purchased an inexpensive tape recorder from an electronics store in Taipei. After using it for several days, you realize that it does not meet your needs. You have brought it back to the store to ask for a refund, but the clerk is insisting that you exchange it for another item of merchandise. Explain to the clerk that the other tape recorders in the store are also inadequate for your purposes and that you do not really need any other type of electronic equipment carried by the store. Explain that you need a cash refund in order to apply the amount to a more sophisticated machine.

On the grounds that proficiency in Chinese at the 3/3+ level requires, in addition to other competencies, the ability to perform appropriately in typical ceremonial/formulaic settings, one of the five role-play questions in each test version dealt with situations similar to the following:

A Chinese colleague in a school in China has invited you and two of your friends to his home for Chinese New Year. You realize that this is a very gracious invitation. After dinner, on behalf of you and your friends, make some appropriate remarks. In the course of your remarks, thank your host and his family for the wonderful meal and for their great hospitality. Also extend an invitation to him and his family for dinner in a local restaurant next Saturday evening.

In addition to insuring that the physical setting, interlocutor(s), temporal or other conditions, and style/tone of the language to be used are clearly indicated in the description of the situation, the following points should be taken into account in drafting and reviewing the "situation" questions:

- The situation should be one in which the examinee could indeed become involved in real life, either as a visitor to a Chinese-speaking country or as a host for Chinese-speaking persons in the United States. As with the "hypothetical" questions previously discussed, the situations should not require the examinee to assume a different persona (for example, a Chinese head-of-household admonishing family members; a Chinese bus driver giving directions to a Chinese passenger, etc.), but should always have real or potential "truth value" for an English-speaking, nonnative learner of Chinese in a realistic Chinese or U.S. context.

- Although the examinee is presumed to have some degree of knowledge of and sensitivity to the Chinese cultural context as this affects the selection and delivery of discourse within the particular situation, only the more basic and pervasive culturally-marked situations should be presented. Cultural situations that a non-native speaker would never or rarely be expected to encounter (e.g., having to say the "right thing" at a Chinese funeral) should be avoided as being both too rare and too specialized.

QUESTION PREPARATION AND REVIEW

For all sections of the test, preparation and review of the test questions is best accomplished by a group of 3-4 individuals who work as a team for initial planning and discussion, subsequently prepare draft questions on an individial basis, and then reconvene to jointly review and revise the draft questions and assemble the draft test form(s) for trial administration. Following are more detailed recommendations for the composition and activities of this group.

(1) The group as a whole should possess educated native-speaker competence in the language; experience in teaching the language to English-speaking learners; educated native-speaker competence in English; and detailed familiarity and considerable "hands-on" experience with the ACTFL/ILR speaking proficiency scale and rating procedure. Ideally, all members of the group would have these qualifications; realistically, these attributes will be shared in different degrees among the group members.

(2) Native-speaker competence and native cultural background in the target language is especially critical in preparing and reviewing the "personal conversation" questions and in reviewing the other sections of the test for cultural authenticity. Native English competence is critical to the final wording of the English directions and English-medium questions/situations.

(3) Two separate, parallel test forms are considered the minimum needed to provide adequate flexibility for multiple administrations (e.g., pre- and post-testing of a given examinee group) and to reduce the likelihood of having portions of the test "memorized" by the students or otherwise compromised within a brief period of time. (The Chinese project developed a total of four forms.) Regardless of the total number of forms to be developed, an adequate "overage" of test questions should be prepared so as to allow for the inevitable attrition at the review stage or following stages. It is not possible to suggest a single percentage figure for "overage," since this would vary according to the technical/linguistic expertise within the group, prior experience in generally similar test development activities, and so forth. However, at least a 50 percent overage (i.e., half again as many draft questions prepared as will be needed in the final test version) would be considered a <u>minimum</u> in this regard.

(4) Draft test questions, after being prepared and reviewed in accordance with the descriptions and discussions in the preceding sections, should be assembled into two or more similar parallel forms for trial administration. Within each test section, the questions should progress in degree of sophistication and anticipated difficulty, so that at the beginning of each section, the examinee will have a "fresh start" with some relatively easy material. (As indicated previously, however, even the more challenging questions should be written in such a way that the less proficient examinee will have at least something to say about the situation or topic, albeit not with the level of sophistication and accuracy expected of examinees at or beyond the level "targeted" by a particular question.)

(5) In addition to the test section-specific considerations, other important points for review are as follows:

(a) Is the target language, wherever used, accurate, current, idiomatic, and of a "standard" variety acceptable to virtually all educated speakers of the language?

(b) Are the pictured settings and verbally defined situations culturally authentic? Do they represent language-use contexts that an English-speaking learner would reasonably be expected to encounter in his or her contacts with native speakers of the language, either in the U.S. or in the target language country?

(c) Are the English directions clear, to the point, and readily understandable by examinees at a level of education and general knowledge anticipated for this test? Although some degree of sophistication and precision of terminology will be required to adequately "set the stage" for certain of the questions (especially in the "responses to printed questions" and "situations" sections), care should be taken to keep the printed and spoken English as simple and as straightforward as possible, consistent with adequately conveying the response tasks at issue.

(d) Do the language-use tasks, considered across the test as a whole, adequately cover the range of topics, functions, and degree of accuracy of examinee performance at issue from levels 1 to 3/3+ on the ACTFL/ILR scale?

ASSEMBLY AND TRIAL ADMINISTRATION OF DRAFT TEST FORMS

Included in this phase are the actual preparation of test booklets and master tape and the small-scale administration of the test to both native speakers of the language and learners representative of the intended examinee population. Procedural suggestions are as follows:

(1) If possible, use an electronic word processing system to prepare both the test booklet copy (other than the line drawings) and the script for the master test tape. This greatly facilitates any necessary later changes, and is also a major timesaver in the preparation of multiple test forms, since uniform "shells" giving the overall format and the unchanging (e.g., general directions) portions of the test can be prepared and used as the basic template for all forms.

(2) For the test booklet, format considerations include the following:

(a) General directions should appear on the front cover, with the examinee instructed not to open the booklet until told to do so by the taped voice. (For the Chinese Speaking Test, the examinee's booklet remains closed until the "simulated conversation" portion has been completed.)

(b) It is advisable to place all textual and graphic material on only the right-hand side of the page, and to have each major test section begin a new page. In this way, maximum control of examinee progess through the test can be maintained, with "reading ahead" and other inappropriate test-taking behavior minimized.

(c) Side-stapled (or preferably, saddle-stapled) booklets are preferable to corner-stapled, in that the former lie flatter and the pages are much less

likely to become detached.

(d) "Do not turn the page until you are asked to do so" (or similar instructions) should be boldly printed at the bottom of each right-hand page.

(e) Paper stock should be sufficiently heavy and opaque to prevent "see-through." Since it is anticipated that each test booklet will serve for multiple administrations (the examinee does not write in or otherwise mark the booklet), heavier paper stock will also lengthen the useful life of the booklet.

(f) For the trial administration, it is sufficient to work with clear, full-size pencil versions of the line drawings (i.e., the drawings that the artist would ultimately ink in for the final version). In this way, any needed changes can be readily and relatively inexpensively incorporated into the final version.

(3) Suggestions for the tape script and for the tape recording session:

(a) In preparation for the trial administration (during which the amount of response time actually taken by typical examinees will be determined), only a short, unobtrusive "beep" signal should be given at the end of each question, followed immediately by the next question. It will be the test administrator's responsibility to stop the master tape after each signal, and then allow the examinee as much time as necessary to respond to that question. In the final version of the test, predetermined response times (i.e., blank tape running for a specified number of seconds) will be inserted following each question as appropriate.

(b) Adequate time (e.g., 5 seconds) should be provided for the examinee to turn the page, as well as 2-3 seconds after a spoken "stop" command. All such timings should be shown in the script at the appropriate locations. Double-spaced scripts are easier to read and handle in a recording studio setting than are single-spaced.

(c) The person voicing the target language questions for the "conversation" portion should be an educated native speaker of the standard dialect with a good recording voice free of regional or idiosyncratic characteristics. Since the conversation on the tape is intended to be with a single speaker, it would be inadvisable to use more than a single voice. Each question should be spoken once, at a relaxed (but <u>not</u> artificially slow) conversational pace.

(d) If at all possible, the English speaker (master voice) should be a trained narrator accustomed to dealing with testing/educational material. At the recording session, qualified listeners capable of carefully monitoring both the English and target language speakers should be on hand, and should not hesitate to call for as many "takes" as necessary to secure a high quality recording.

(4) Once trial versions of the test booklet and tape are available (both containing somewhat more questions in each section than will ultimately be included in the final test form), the test should be administered to 3-4 educated native speakers of the target language on a one-to-one basis. At

issue in this tryout are:

(a) Accuracy and adequacy of the stimulus questions. Do the test questions in each section, including the situation descriptions, elicit the types of responses anticipated? Do the higher-level questions yield native-speaker answers at the expected level of sophistication, discourse style, etc.?

(b) Are there ambiguities in any aspect of the line drawings? For the picture descriptions and sequence of pictures, do the native speakers "follow the script" as to the anticipated responses?

(c) Do any of the test questions or other aspects of the test strike the native speakers as linguistically unnatural, culturally questionable, etc.?

(d) Do the native speakers take roughly similar amounts of time to answer a given question? The typical native response time may be considered a minimum, which will usually need to be increased for non-native examinees (see discussion below).

(5) Following any revisions indicated on the basis of the native-speaker tryouts, the draft test materials should then be administered--again on a one-to-one basis-- to 8-10 examinees distributed as evenly as possible over the level 1 to level 3/3+ range. The following are especially to be noted:

(a) Under "natural" (non-timed) conditions, what is the range of response times for each of the test questions? It should be possible, for each question, to identify a general range of response times within which most of the examinees appear to fall, discounting, as necessary, abnormally brief responses by examinees at the lowest end of the scale and/or an occasional quite garrulous reply by a more fluent examinee. The response time finally established for a given question should tend toward the upper bound of the identified range. As a rough estimate of suitable response times, which should be experimentally verified before being adopted for a given test and language, the following are the timings used in the Chinese test project:

Simulated Conversation:	5-20 sec.
Picture Descriptions:	75 sec.
Responses to Printed Questions:	75-90 sec.
Situations:	30-90 sec.

(b) Do any of the examinees seem to go "off track" in their responses (i.e., give or attempt to give answers that are not at all in keeping with the intent of the question)? If so, the examinee should be asked to comment on the thought processes he or she went through in analyzing and responding to the question, with possible implications for revision.

(c) Structured conversation with the examinees immediately following the testing can help provide information on unclear questions, portions of the test that the examinee considered "unfair," etc. Since the test is proficiency-oriented, inordinate attention should not be paid to the fact that a given examinee "never learned" to handle a particular topic, style of discourse, or

general lexical area. However, if several otherwise fluent examinees appear to have "blocked" on the word for a particular object or action in one of the drawings, or in some other way been appreciably hindered by the lack of a particular lexical item, consideration should be given to revising the question to minimize or eliminate this particular stumbling block.

FINAL TEST PRODUCTION AND ADMINISTRATION

In view of the fact that the tryout version of the test will contain somewhat more questions than are to be used in the final version, it will be possible at this point to remove from further consideration questions that appear to have quality deficiencies of any type and to concentrate on those questions which have been found to operate substantially in the manner intended. Any necessary corrections or modifications should, of course, be made to the retained questions with respect to both the printed text/pictures and the recorded script. Some judgment will be required concerning whether the tape script should be re-recorded in its entirety or whether it will be possible to simply "strip in" any needed corrections on the original master tape. As a general matter, unless the changes are quite minor and the services of a professional recording studio are available to insure, for example, that the recording level of the inserted material closely matches the original level, it is preferable to re-record the entire test tape, with, of course, the same careful during- and after-recording review that was carried out for the trial version.

Once the proof copy of the final master tape is available and the camera-ready copy for the test booklet has been prepared (with the latter mocked-up into the exact format and pagination that will be used in the printed version), at least one--and preferably more--qualified reviewers should attentively "take" the test, from beginning to end, by listening to the tape and following all of the directions provided. It is at this stage that any remaining problems will need to be caught (for example, insufficient time for page turning, mis-numbered pages, etc.).

If possible, offset printing, rather than xeroxing or other similar process, should be used to produce the test booklets. Particularly for the line drawings, offset-quality printing is necessary to provide highly legible and unambigous reproduction. As previously indicated, opaque, somewhat heavier than normal paper stock is recommended, especially if the booklets are to be reused several times. Side- or saddle-stapling is recommended so that the pages will lie flatter and the booklet will be more physically secure than with corner stapling.

The completed test is readily administerable either in a language laboratory equipped for student recording or, on an examinee-by-examinee basis, by means of two portable tape recorders, one to play the tape and the other to record the examinee's voice. For language lab administration, examinees should be seated as far from one another as possible, to minimize deliberate or inadvertent "listening in" to other examinees' responses.

TEST SCORING

As emphasized at the beginning of this manual, the conceptualization and development of the particular type of test described here is aimed at implementing, in a semi-direct format, the proficiency testing approach and rating scale developed by ACTFL/ILR and other agencies collaborating in these efforts. It is, therefore, considered absolutely necessary for persons attempting to operationally evaluate examinee performance by means of a test of this type to have undergone a full interviewer/rater training program under the auspices of ACTFL or other recognized agency, so as to be able to validly and reliably apply the same evaluation process and standards to the semi-direct instrument. Information concerning the availability and scheduling of interviewer/rater training workshops conducted by ACTFL or other qualified training sources may be obtained by contacting the American Council on the Teaching of Foreign Languages, 579 Broadway, Hastings-on-Hudson, NY 10706.

Experience gained during the test scoring and validation phases of the Chinese Speaking Test project suggests that linguistically qualified individuals who have successfully completed a comprehensive ACTFL-type interviewer/rater training program should be able to readily transfer these evaluation standards and skills to similar assessments of examinee performance based on responses to the semi-direct test. Since the latter type of testing work is still in its early stages, there has not yet been the opportunity to develop a large infrastructure of auxiliary information and shared experience particular to the rating of the semi-direct tests per se. However, it is CAL's intent to serve as an informal clearinghouse for information, suggestions, and accounts-of-experience with the development and evaluation of both the present semi-direct tests in Chinese and similar tests in other languages that may be developed in the future. Persons or institutions who would like to receive or share any relevant information are asked to contact the Center for Applied Linguistics, 1118 22nd Street, NW, Washington, DC 20037.

References:

Clark, John L. D. "Direct vs. Semi-Direct Tests of Speaking Proficiency," pp. 35-49 in Eugene J. Briere and Frances B. Hinofotis, eds., Concepts in Language Testing: Some Recent Studies. Washington, DC: Teachers of English to Speakers of Other Languages.

Lowe, Pardee, Jr. and Ray T. Clifford. "Developing an Indirect Measure of Overall Oral Proficiency," pp. 31-39 in James R. Frith, ed., Measuring Spoken Language Proficiency. Washington, DC: Georgetown University Press, 1980.

Questionnaire for Participants in Chinese Speaking Test Study

We would very much appreciate your answers to the following questions concerning the three speaking tests (live interview, plus two tape-recorded tests) which you took in connection with the Chinese Speaking Test study. We are asking for your name on the form simply as a check that the questionnaire has been received. All information will be analyzed on an anonymous, group basis for purposes of test improvement. Thank you in advance for your comments, and for your much-appreciated participation in the test development process.

1. Over the course of the live interview, do you feel that your maximum level of speaking ability in Chinese was adequately probed by the tester?

( ) Yes ( ) No

Any comments on this question?

2. Over the course of the taped test, do you feel that the descriptions, narratives, situations, and other types of questions in the test were adequate to probe your maximum level of speaking ability in Chinese?

( ) Yes ( ) No

Comment on this question?

3. In the live interview, were there any questions asked or speaking situations required which you felt were in any way "unfair"?

( ) Yes

( ) No

(If "Yes," please briefly describe the question(s) involved.)

4. In the taped tests, were there any picture/descriptions, narratives, situations, or other questions that you felt were in any way "unfair"?

( ) Yes ( ) No

(If "Yes," please describe briefly.)

5. In which of the two types of test--live interview or taped test--did you feel the more anxious or nervous?

( ) Live interview

( ) Taped test

( ) Both were about the same

Comments?

6. Which of the two types of tests did you feel was the more difficult?

( ) Live interview

( ) Taped test

( ) Both were about the same

Comments?

7. During the taped test, were the pauses for your response usually long enough for you to respond as fully as you wished (or were able)?

( ) Pauses were usually too short

( ) Pauses were usually about right

( ) Pauses were usually too long

Comments?

8. In the taped test, were the directions for the different parts of the test sufficiently clear and detailed?

( ) Yes ( ) No

Comments?

9. Assuming that you would receive the same score through both techniques, would you personally rather take a live interview or a (single) taped test in order to show your speaking ability?

( ) I would rather take a live interview because:

( ) I would rather take a taped test because:

( ) I have no real preference

10. Please use the space below to comment on any aspect of the live interview or the taped test that is not covered in any of the preceding questions. We would appreciate any suggestions as to how these testing procedures might be improved.

Thank you very much for your help.

Name _____